



(19)

Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 0 843 848 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:
07.04.2004 Bulletin 2004/15

(51) Int Cl.7: **G06F 9/30, G06F 9/318,
G06F 9/38**

(21) Application number: **97918307.6**

(86) International application number:
PCT/IB1997/000558

(22) Date of filing: **15.05.1997**

(87) International publication number:
WO 1997/043710 (20.11.1997 Gazette 1997/50)

(54) VLIW PROCESSOR WHICH PROCESSES COMPRESSED INSTRUCTION FORMAT

VLIW-PROZESSOR ZUR VERARBEITUNG EINES KOMPRIMIERTEN INSTRUKTIONSFOMATS
PROCESSEUR VLIW TRAITANT DES FORMATS D'INSTRUCTIONS COMPRIMEES

(84) Designated Contracting States:
DE FR GB IE IT NL

- **HAMPAPURAM, Hari
5656 AA Eindhoven (NL)**
- **LEE, Yen, C.
5656 AA Eindhoven (NL)**

(30) Priority: **15.05.1996 US 648359
15.05.1996 US 649733
15.05.1996 US 649731**

(74) Representative: **de Jong, Durk Jan et al
Philips
Intellectual Property & Standards
P.O. Box 220
5600 AE Eindhoven (NL)**

(43) Date of publication of application:
27.05.1998 Bulletin 1998/22

(56) References cited:
WO-A-95/19006 US-A- 5 179 680

(73) Proprietor: **Koninklijke Philips Electronics N.V.
5621 BA Eindhoven (NL)**

- **MICROPROCESSOR REPORT, Vol. 8, No. 16,
p12(4), 5 Dec. 1994, "Philips Hopes to Displace
DSPs with VLIW: TriMedia Processors Aimed at
Future Multimedia Embedded Apps. Brian
Case".**

(72) Inventors:

- **JACOBS, Eino
5656 AA Eindhoven (NL)**
- **ANG, Michael
5656 AA Eindhoven (NL)**

EP 0 843 848 B1

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

[0001] The invention relates to VLIW (Very Long Instruction Word) processors and in particular to instruction formats for such processors and an apparatus and method for processing such instruction formats.

5 [0002] The invention relates in particular to a VLIW processor according to the part of Claim 1 preceding the words "characterized in that".

[0003] A scheme for compression of VLIW instructions has been proposed in US Pat. No.s 5,179,680 and 5,057,837. This compression scheme provides for an instruction word from which unused operations are eliminated and a mask word indicating which operations have been eliminated.

10 [0004] VLIW processors have instruction words including a plurality of issue slots. The processors also include a plurality of functional units. Each functional unit is for executing a set of operations of a given type. Each functional unit is RISC-like in that it can begin an instruction in each machine cycle in a pipe-lined manner. Each issue slot is for holding a respective operation. All of the operations in a same instruction word are to be begun in parallel on the functional unit in a single cycle of the processor. Thus the VLIW implements fine-grained parallelism.

15 [0005] Thus, typically an instruction on a VLIW machine includes a plurality of operations. On conventional machines, each operation might be referred to as a separate instruction. However, in the VLIW machine, each instruction is composed of operations or no-ops (dummy operations).

20 [0006] Like conventional processors, VLIW processors use a memory device, such as a disk drive to store instruction streams for execution on the processor. A VLIW processor can also use caches, like conventional processors, to store pieces of the instruction streams with high bandwidth accessibility to the processor.

25 [0007] The instruction in the VLIW machine is built up by a programmer or compiler out of these operations. Thus the scheduling in the VLIW processor is software-controlled.

[0008] The VLIW processor can be compared with other types of parallel processors such as vector processors and superscalar processors as follows. Vector processors have single operations which are performed on multiple data items simultaneously. Superscalar processors implement fine-grained parallelism, like the VLIW processors, but unlike the VLIW processor, the superscalar processor schedules operations in hardware.

30 [0009] Because of the long instruction words, the VLIW processor has aggravated problems with cache use. In particular, large code size causes cache misses, i.e. situations where needed instructions are not in cache. Large code size also requires a higher main memory bandwidth to transfer code from the main memory to the cache.

[0010] Large code size can be aggravated by the following factors.

- In order to fine tune programs for optimal running, techniques such as grafting, loop unrolling, and procedure inlining are used. These procedures increase code size.
- Not all issue slots are used in each instruction. A good optimizing compiler can reduce the number of unused issue slots; however a certain number of no-ops (dummy instructions) will continue to be present in most instruction streams.
- In order to optimize use of the functional units, operations on conditional branches are typically begun prior to expiration of the branch delay, i.e. before it is known which branch is going to be taken. To resolve which results are actually to be used, guard bits are included with the instructions.
- Larger register files, preferably used on newer processor types, require longer addresses, which have to be included with operations.

35 [0011] A scheme for compression of VLIW instructions has been proposed in US Pat. No.s 5,179,680 and 5,057,837. This compression scheme eliminates unused operations in an instruction word using a mask word, but there is more room to compress the instruction.

40 [0012] Further information about technical background to this application can be found in the following prior applications:

- US Application Ser. No. 998,090, filed December 29, 1992 (PHA 21,777), which shows a VLIW processor architecture for implementing fine-grained parallelism;
- US Application Ser. No. 142,648 filed October 25, 1993 (PHA 1205) (US-A-5,450,556), which shows use of guard bits; and
- US Application Ser. No. 366,958 filed December 30, 1994 (PHA 21,932) (EP-A-0 748 477) which shows a register file for use with VLIW architecture.

45

Bibliography of program compression techniques:

- J. Wang et al, "The Feasibility of Using Compression to Increase Memory System Performance", Proc. 2nd Int.

Workshop on Modeling Analysis, and Simulation of Computer and Telecommunications Systems, p. 107-113 (Durham, NC, USA 1994);

- H. Schröder et al., "Program compression on the instruction systolic array", Parallel Computing, vol. 17, n 2-3, June 1991, p.207-219;
- 5 - A. Wolfe et al., "Executing Compressed Programs on an Embedded RISC Architecture", J. Computer and Software Engineering, vol. 2, no. 3, pp 315-27, (1994);
- M. Kozuch et al., "Compression of Embedded Systems Programs", Proc. 1994 IEEE Int. Conf. on Computer Design: VLSI in Computers and Processors (Oct. 10-12, 1994, Cambridge MA, USA) pp.270-7.

10 Typically the approach adopted in these documents has been to attempt to compress a program as a whole or blocks of program code. Moreover, typically some table of instruction locations or locations of blocks of instructions is necessitated by these approaches.

[0013] It is an object of the invention to reduce code size in a VLIW processor.

15 [0014] It is another object of the invention to create a VLIW processor which processes more highly compressed instructions.

20 [0015] The processor according to the invention is characterized in that the decompression unit is arranged to decompress operations with respective compressed operation lengths chosen from a plurality of finite lengths, which finite lengths include at least two non-zero lengths. The set of available operation lengths is for example 0, 26, 34 and 42 bit long compressed operations. Which operations are compressed to a particular length depends first of all on a study of frequency of occurrence of the operations. This could vary depending on the type of software written. Furthermore the length may be made dependent on whether the operation is guarded or unguarded, whether it produces a result, whether it uses an immediate parameter and on the number of operands it uses.

25 [0016] The processor according to the invention has an embodiment wherein the decompression unit is arranged to take a format field from the compressed instruction medium, the format field specifying the respective compressed operation length for each operation of the compressed instruction, the decompression unit decompressing the operations of the compressed instruction according to the format field. Preferably the format field also specifies which issue slots of the processor are to be used by the instruction.

30 [0017] In a further embodiment of the processor according to the invention the format field has N sub-fields, N being the number of issue slots, each sub-field specifying a compressed operation length for a respective issue slot, characterized in that the sub-fields each contain at least two bits. When four different operation lengths are used, the sub-field may be for example 2-bit long.

[0018] In another embodiment of the invention wherein the decompression unit is arranged for

- 35 - taking a preceding compressed instruction from the compressed instruction medium together with the format field,
- starting decompression of the preceding compressed instruction and subsequently
- taking the compressed instruction from the compressed instruction memory and starting decompression of the compressed instruction according to the format field taken from the compressed instruction medium together with the preceding compressed instruction. Thus the format field is available before the compressed instruction is loaded and preparations for decompression according to the format field can start before the compressed instruction is loaded.

40 [0019] In an embodiment of the invention, the compression unit takes the format field from the compressed instruction medium in a memory access unit, the memory access unit also comprising at least one operation part sub-field, the decompression unit integrating the operation part sub-field in at least one of the operations of the decompressed instruction. This increases retrieval efficiency and allows pipelining of instruction retrieval. This may be used in a processor capable of decompressing any number of available operation lengths, also two lengths (e.g. 0 and 32). Thus, the decompression unit is alerted to how many issue slots to expect in the instruction to follow prior to retrieval of that instruction. For each instruction, other than branch targets, a field specifying a format may be stored with a previous instruction.

50 [0020] The invention also relates to a method of producing compressed code for running on a VLIW processor according to Claim 7. This method generates instructions useful for the processor.

55 [0021] The method according to the invention has an embodiment in which the method is applied to a stream of instructions including said instruction, the method comprising the step of determining for each instruction whether that instruction is a branch target of a branch from another instruction of the stream of instructions, and compressing only those instructions which are not branch targets. Thus branch target will not need to be decompressed during execution of a program by the processor and no delay to decompress is needed after execution of a branch instruction.

BRIEF DESCRIPTION OF THE DRAWING

[0022] The invention will now be described by way of non-limitative example with reference to the following figures:

- 5 Fig. 1a shows a processor for using the compressed instruction format of the invention.
Fig. 1b shows more detail of the CPU of the processor of Fig. 1a.
Figs. 2a-2e show possible positions of instructions in cache.
Fig. 3 illustrates a part of the compression scheme in accordance with the invention.
Figs. 4a - 4f illustrate examples of compressed instructions in accordance with the invention.
10 Figs. 5a-5b give a table of compressed instructions formats according to the invention.
Fig. 6a is a schematic showing the functioning of instruction cache 103 on input.
Fig. 6b is a schematic showing the functioning of a portion of the instruction cache 103 on output.
Fig. 7 is a schematic showing the functioning of instruction cache 104 on output.
Fig. 8 illustrates compilation and linking of code according to the invention.
15 Fig. 9 is a flow chart of compression and shuffling modules.
Fig. 10 expands box 902 of Fig. 9.
Fig. 11 expands box 1005 of Fig. 10.
Fig. 12 illustrates the decompression process.

20 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0023] Fig. 1a shows the general structure of a processor according to the invention. A microprocessor according to the invention includes a CPU 102, an instruction cache 103, and a data cache 105. The CPU is connected to the caches by high bandwidth buses. The microprocessor also contains a memory 104 where an instruction stream is stored.

[0024] The cache 103 is structured to have 512 bit double words. The individual bytes in the words are addressable, but the bits are not. Bytes are 8 bits long. Preferably the double words are accessible as a single word in a single clock cycle.

[0025] The instruction stream is stored as instructions in a compressed format in accordance with the invention. The compressed format is used both in the memory 104 and in the cache 103.

[0026] Fig. 1b shows more detail of the VLIW processor according to the invention. The processor includes a multiport register file 150, a number of functional units 151, 152, 153, ..., and an instruction issue register 154. The multiport register file stores results from and operands for the functional units. The instruction issue register includes a plurality of issue slots for containing operations to be commenced in a single clock cycle, in parallel, on the functional units 151, 152, 153, A decompression unit 155, explained more fully below, converts the compressed instructions from the instruction cache 103 into a form usable by the IIR 154.

COMPRESSED INSTRUCTION FORMAT

40 1. General Characteristics

[0027] The preferred embodiment of the claimed instruction format is optimized for use in a VLIW machine having an instruction word which contains 5 issue slots. The format has the following characteristics

- 45 - unaligned, variable length instructions;
- variable number of operations per instruction;
- 3 possible sizes of operations: 26, 34 or 42 bits (also called a 26/34/42 format).
- the 32 most frequently used operations are encoded more compactly than the other operations;
- operations can be guarded or unguarded;
50 - operations are one of zeroary, unary, or binary, i.e. they have 0, 1 or 2 operands;
- operations can be resultless;
- operations can contain immediate parameters having 7 or 32 bits
- branch targets are not compressed; and
- format bits for an instruction are located in the prior instruction.

55 2. Instruction Alignment

[0028] Except for branch targets, instructions are stored aligned on byte boundaries in cache and main memory.

Instructions are unaligned with respect to word or block boundaries in either cache or main memory. Unaligned instruction cache access is therefore needed.

[0029] In order to retrieve unaligned instructions, processor retrieves one word per clock cycle from the cache.

[0030] As will be seen from the compression format described below, branch targets need to be uncompressed and must fall within a single word of the cache, so that they can be retrieved in a single clock cycle. Branch targets are aligned by the compiler or programmer according to the following rule:

if a word boundary falls within the branch target or exactly at the end of the branch target, padding is added to make the branch target start at the next word boundary

Because the preferred cache retrieves double words in a single clock cycle, the rule above can be modified to substitute double word boundaries for word boundaries.

[0031] The normal unaligned instructions are retrieved so that succeeding instructions are assembled from the tail portion of the current word and an initial portion of the succeeding word. Similarly, all subsequent instructions may be assembled from 2 cache words, retrieving an additional word in each clock cycle.

[0032] This means that whenever code segments are relocated (for instance in the linker or in the loader) alignment must be maintained. This can be achieved by relocating base addresses of the code segments to multiples of the cache block size.

[0033] Figs. 2a-e show unaligned instruction storage in cache in accordance with the invention.

[0034] Fig. 2a shows two cache words with three instructions i1, i2, and i3 in accordance with the invention. The instructions are unaligned with respect to word boundaries. Instructions i1 and i2 can be branch targets, because they fall entirely within a cache word. Instruction i3 crosses a word boundary and therefore must not be a branch target. For the purposes of these examples, however, it will be assumed that i1 and only i1 is a branch target.

[0035] Fig. 2b shows an impermissible situation. Branch target i1 crosses a word boundary. Accordingly, the compiler or programmer must shift the instruction i1 to a word boundary and fill the open area with padding bytes, as shown in Fig. 2c.

[0036] Fig. 2d shows another impermissible situation. Branch target instruction i1 ends precisely at a word boundary. In this situation, again i1 must be moved over to a word boundary and the open area filled with padding as shown in Fig. 2e.

[0037] Branch targets must be instructions, rather than operations within instructions. The instruction compression techniques described below generally eliminate no-ops (dummy instructions). However, because the branch target instructions are uncompressed, they must contain no-ops to fill the issue slots which are not to be used by the processor.

3. Bit and Byte order

[0038] Throughout this application bit and byte order are little endian. Bits and bytes are listed with the least significant bits first, as below:

	Bit number 0....8....16....		
40	Byte number	0	1 2
	address	0	1 2

4. Instruction format

[0039] The compressed instruction can have up to seven types of fields. These are listed below. The format bits are the only mandatory field.

[0040] The instructions are composed of byte aligned sections. The first two bytes contain the format bits and the first group of 2-bit operation parts. All of the other fields are integral multiples of a byte, except for the second 2-bit operation parts which contain padding bits.

[0041] The operations, as explained above can have 26, 34, or 42 bits. 26-bit operations are broken up into a 2-bit part to be stored with the format bits and a 24-bit part. 34-bit operations are broken up into a 2 bit part, a 24-bit part, and a one byte extension. 42-bit operations are broken up into a 2 bit part, a 24 bit part, and a two byte extension.

A. Format bits

[0042] These are described in section 5 below. With a 5 issue slot machine, 10 format bits are needed. Thus, one byte plus two bits are used.

B. 2-bit operation parts, first group

[0043] While most of each operation is stored in the 24-bit part explained below, i.e. 3 bytes, with the preferred instruction set 24 bits was not adequate. The shortest operations required 26 bits. Accordingly, it was found that the six bits left over in the bytes for the format bit field could advantageously be used to store extra bits from the operations, two bits for each of three operations. If the six bits designated for the 2-bit parts are not needed, they can be filled with padding bits.

C. 24-bit operation parts, first group

[0044] There will be as many 24 bit operation parts as there were 2 bit operation parts in the two bit operation parts, first group. In other words, up to three 3 byte operation parts can be stored here.

D. 2 bit operation parts, second group

[0045] In machines with more than 3 issue slots a second group of 2-bit and 24-bit operation parts is necessary. The second group of 2-bit parts consists of a byte with 4 sets of 2-bit parts. If any issue slot is unused, its bit positions are filled with padding bits. Padding bits sit on the left side of the byte. In a five issue slot machine, with all slots used, this section would contain 4 padding bits followed by two groups of 2-bit parts. The five issue slots are spread out over the two groups: 3 issue slots in the first group and 2 issue slots in the second group.

E. 24-bit operation parts, second group

[0046] The group of 2-bit parts is followed by a corresponding group of 24 bit operation parts. In a five issue slot machine with all slots used, there would be two 24-bit parts in this group.

F. further groups of 2-bit and 24-bit parts

[0047] In a very wide machine, i.e. more than 6 issue slots, further groups of 2-bit and 24-bit operation parts are necessary.

G. Operation extension

[0048] At the end of the instruction there is a byte-aligned group of optional 8 or 16 bit operation extensions, each of them byte aligned. The extensions are used to extend the size of the operations from the basic 26 bit to 34 or 42 bit, if needed.

[0049] The formal specification for the instruction format is:

40

45

50

55

```

<instruction> ::=

    <instruction start>
5
    <instruction middle>
    <instruction end>
    <instruction extension>

10   <instruction start> ::=

        <Format:2*N>{ <padding:1>}V2{ <2-bit operation part:2>}V1{ <24-
            bit operation part:24>}V1

15   <instruction middle> ::= {{ <2-bit operation part:2>}4 {24-bit      operation
            part:24>}4}V3

20   <instruction end> ::= { <padding:1>}V5{ <2-bit operation  part:2>}V4      {24-bit
            operation part:24>}V4

25   <instruction extension> ::= { <operationextension:0/8/16>}S
    <padding> ::= "0"

```

Wherein the variables used above are defined as follows:

```

30   N = the number of issue slots of the machine, N>1
       S = the number of issue slots used in this instruction (0≤S≤N)
       C1 = 4 - (N mod 4)
       If (S ≤ C1) then V1=S and V2 = 2*(C1-V1)
       If (S > C1) then V1=C1 and V2 =0
       V3 = (S-V1) div 4
35   V4 = (S-V1) mod 4
       If (V4 > 0) then V5 = 2*(4-V4) else V5=0

```

40

45

50

55

Explanation of notation

[0050]

5	::=	means	"is defined as"
	< field name:number >		
10		means	the field indicated before the colon has the number of bits indicated after the colon.
15	{ < field name > }number	means	the field indicated in the angle brackets and braces is repeated the number of times indicated after the braces
20	"0"	means	the bit "0".
25	"div"	means	integer divide
	"mod"	means	modulo
30	:0/8/16	means	that the field is 0, 8, or 16 bits long

[0051] Examples of compressed instructions are shown in Figs. 4 a-f.

[0052] Fig. 4a shows an instruction with no operations. The instruction contains two bytes, including 10 bits for the format field and 6 bits which contain only padding. The former is present in all the instructions. The latter normally correspond to the 2-bit operation parts. The X's at the top of the bit field indicate that the fields contain padding. In the later figures, an O is used to indicate that the fields are used.

[0053] Fig. 4b shows an instruction with one 26-bit operation. The operation includes one 24 bit part at bytes 3-5 and one 2 bit part in byte 2. The 2 bits which are used are marked with an O at the top.

[0054] Fig. 4c shows an instruction with two 26-bit operations. The first 26-bit operation has its 24-bit part in bytes 3-5 and its extra two bits in the last of the 2-bit part fields. The second 26-bit operation has its 24-bit part in bytes 6-8 and its extra two bits in the second to last of the 2-bit part fields.

[0055] Fig. 4d shows an instruction with three 26-bit operations. The 24-bit parts are located in bytes 3-11 and the 2-bit parts are located in byte 2 in reversed order from the 24-bit parts.

[0056] Fig. 4e shows an instruction with four operations. The second operation has a 2 byte extension. The fourth operation has a one byte extension. The 24-bit parts of the operations are stored in bytes 3-11 and 13-15. The 2-bit parts of the first three operations are located in byte 2. The 2-bit part of the fourth operation is located in byte 12. An extension for operation 2 is located in bytes 16-17. An extension for operation 4 is located in byte 18.

[0057] Fig. 4f shows an instruction with 5 operations each of which has a one byte extension. The extensions all appear at the end of the instruction.

[0058] While extensions only appear after the second group of 2-bit parts in the examples, they could equally well appear at the end of an instruction with 3 or less operations. In such a case the second group of 2-bit parts would not be needed.

[0059] There is no fixed relationship between the position of operations in the instruction and the issue slot in which they are issued. This makes it possible to make an instruction shorter when not all issue slots are used. Operation positions are filled from left to right. The Format section of the instruction indicates to which issue slot a particular operation belongs. For instance, if any instruction contains only one operation, then it is located in the first operation position and it can be issued to any issue slot, not just slot number 1. The decompression hardware takes care of routing operation to their proper issue slots.

[0060] No padding bytes are allowed between instructions that form one sequential block of code. Padding blocks are allowed between distinct blocks of code.

5. Format Bits

[0061] The instruction compression technique of the invention is characterized by the use of a format field which specifies which issue slots are to be used by the compressed instruction. To achieve retrieval efficiency, format bits are stored in the instruction preceding the instruction to which the format bits relate. This allows pipelining of instruction retrieval. The decompression unit is alerted to how many issue slots to expect in the instruction to follow prior to retrieval of that instruction. The storage of format bits preceding the operations to which they relate is illustrated in Fig. 3. Instruction 1, which is an uncompressed branch target, contains a format field which indicates the issue slots used by the operations specified in instruction 2. Instructions 2 through 4 are compressed. Each contains a format field which specifies issue slots to be used by the operations of the subsequent instruction.

[0062] The format bits are encoded as follows. There are 2^*N format bits for an N-issue slot machine. In the case of the preferred embodiment, there are five issue slots. Accordingly, there are 10 format bits. Herein the format bits will be referred to in matrix notation as $\text{Format}[j]$ where j is the bit number. The format bits are organized in N groups of 2 bits. Bits $\text{Format}[2i]$ and $\text{Format}[2i+1]$ give format information about issue slot i , where $0 \leq i \leq N$. The meaning of the format bits is explained in the following table:

TABLE I

Format [2i] lsb	Format[2i+1] msb	meaning
0	0	Issue slot i is used and an operation for it is available in the instruction. The operation size is 26 bits. The size of the extension is 0 bytes
1	0	Issue slot i is used and an operation for it is available in the instruction. The operation size is 34 bits. The size of the extension is 1 byte.
0	1	Issue slot i is used and an operation for it is available in the instruction. The operation size is 42 bits. The size of the extension is 2 bytes.
1	1	Issue slot i is unused and no operation for it is included in the instruction.

[0063] Operations correspond to issue slots in left to right order. For instance, if 2 issue slots are used, and $\text{Format} = \{1, 0, 1, 1, 1, 0, 1, 1\}$, then the instruction contains two 34 bit operations. The left most is routed to issue slot 0 and the right most is routed to issue slot 3. If $\text{Format} = \{1, 1, 1, 1, 1, 0, 1, 0, 1, 0\}$, then the instruction contains three 34 bit operations, the left most is routed to issue slot 2, the second operation is intended for issue slot 3, and the right most belongs to issue slot 4.

[0064] The format used to decompress branch target instructions is a constant. $\text{Constant_Format} = \{0, 1, 0, 1, 0, 1, 0, 1, 0, 1\}$ for the preferred five issue slot machine.

6. Operation Formats

[0065] The format of an operation depends on the following properties

- zeroary, unary, or binary;
- parametric or non-parametric. Parametric instructions contain an immediate operand in the code. Parameters can be of differing sizes. Here there are param7, i.e. seven bit parameters, and param32, i.e. 32 bit parameters.
- result producing or resultless;
- long or short op code. The short op codes are the 32 most frequent op codes and are five bits long. The long op codes are eight bits long and include all of the op codes, including the ones which can be expressed in a short format. Op codes 0 to 31 are reserved for the 32 short op codes
- guarded or unguarded. An unguarded instruction has a constant value of the guard of TRUE.
- latency. A format bit indicates if operations have latency equal to one or latency larger than 1.
- signed/unsigned. A format bit indicates for parametric operations if the parameter is signed or unsigned.

[0066] The guarded or unguarded property is determined in the uncompressed instruction format by using the special register file address of the constant 1. If a guard address field contains the address of the constant 1, then the operation

is unguarded, otherwise it is guarded. Most operations can occur both in guarded and unguarded formats. An immediate operation, i.e. an operation which transfers a constant to a register, has no guard field and is always unguarded.

[0067] Which op codes are included in the list of 32 short op codes depends on a study of frequency of occurrence which could vary depending on the type of software written.

[0068] The table II below lists operation formats used by the invention. Unless otherwise stated, all formats are: not parametric, with result, guarded, and long op code. To keep the tables and figures as simple as possible the following table does not list a special form for latency and signed/unsigned properties. These are indicated with L and S in the format descriptions. For non-parametric, zeroary operations, the unary format is used. In that case the field for the argument is undefined.

10

TABLE II

OPERATION TYPE	SIZE
< binary-unguarded-short >	26
< unary-param7-unguarded-short >	26
< binary-unguarded-param7-resultless-short >	26
< unary-short >	26
< binary-short >	34
< unary-param7-short >	34
< binary-param7-resultless-short >	34
< binary-unguarded >	34
< binary-resultless >	34
< unary-param7-unguarded >	34
< unary >	34
< binary-param7-resultless >	42
< binary >	42
< unary-param7 >	42
< zeroary-param32 >	42
< zeroary-param32-resultless >	42

[0069] For all operations a 42-bit format is available for use in branch targets. For unary and binary-resultless operations, the < binary> format can be used. In that case, unused fields in the binary format have undefined values. Short 5-bit op codes are converted to long 8-bit op codes by padding the most significant bits with 0's. Unguarded operations get as a guard address value, the register file address of constant TRUE. For store operations the 42 bit, binary-param7-resultless> format is used instead of the regular 34 bit < binary-param7-resultless short> format (assuming store operations belong to the set of short operations).

[0070] Operation types which do not appear in table II are mapped onto those appearing in table II, according to the following table of aliases:

45

TABLE II'

FORMAT	ALIASED TO
zeroary	unary
unary_resultless	unary
binary_resultless_short	binary_resultless
zeroary_param32_short	zeroary_param32
zeroary_param32_resultless_short	zeroary_param32_resultless
zeroary_short	unary

EP 0 843 848 B1

TABLE II' (continued)

FORMAT	ALIASED TO
unary_resultless_short	unary
binary_resultless_unguarded	binary_resultless
unary_unguarded	unary
binary_param7_resultless_unguarded	binary_param7_resultless
unary_unguarded	unary
binary_param7_resultless_unguarded	binary_param7_resultless
zeroary_unguarded	unary
unary_resultless_unguarded_short	binary_unguarded_short
unary_unguarded_short	unary_short
zeroary_param32_unguarded_short	zeroary_param32
zeroary_param32_resultless_unguarded_short	zeroary_param32_resultless
zeroary_unguarded_short	unary
unary_resultless_unguarded_short	unary
unary_long	binary
binary_long	binary
binary_resultless_long	binary
unary_param7_long	unary_param7
binary_param7_resultless_long	binary_param7_resultless
zeroary_param32_long	zeroary_param32
zeroary_param32_resultless_long	zeroary_param32_resultless
zeroary_long	binary
unary_resultless_long	binary

[0071] The following is a table of fields which appear in operations:

TABLE III

FIELD	SIZE	MEANING
src1	7	register file address of first operand
src2	7	register file address of second operand
guard	7	register file address of guard
dst	7	register file address of result
param	7/32	7 bit parameter or 32 bit immediate value
op code	5/8	5 bit short op code or 8 bit long op code

[0072] Fig. 5 includes a complete specification of the encoding of operations.

7. Extensions of the instruction format

[0073] Within the instruction format there is some flexibility to add new operations and operation forms, as long as encoding within a maximum size of 42 bits is possible.

[0074] The format is based on 7-bit register file address. For register file addresses of different sizes, redesign of the format and decompression hardware is necessary.

[0075] The format can be used on machines with varying numbers of issue slots. However, the maximum size of the instruction is constrained by the word size in the instruction cache. In a 4 issue slot machine the maximum instruction size is 22 bytes (176 bits) using four 42-bit operations plus 8 format bits. In a five issue slot machine, the maximum instruction size is 28 bytes (224 bits) using five 42-bit operations plus 10 format bits.

[0076] In a six issue slot machine, the maximum instruction size would be 264 bits, using six 42-bit operations plus 12 format bits. If the word size is limited to 256 bits, and six issue slots are desired, the scheduler can be constrained to use at most 5 operations of the 42 bit format in one instruction. The fixed format for branch targets would have to use 5 issue slots of 42 bits and one issue slot of 34 bits.

10 COMPRESSING THE INSTRUCTIONS

[0077] Fig. 8 shows a diagram of how source code becomes a loadable, compressed object module. First the source code 801 must be compiled by compiler 802 to create a first set of object modules 803. These modules are linked by linker 804 to create a second type of object module 805. This module is then compressed and shuffled at 806 to yield loadable module 807.

Any standard compiler or linker can be used. Object modules II contain a number of standard data structures. These include: a header; global & local symbol tables; reference table for relocation information; a section table; and debug information, some of which are used by the compression and shuffling module 807. The object module II also has partitions, including a text partition, where the instructions to be processed reside, and a source partition which keeps track of which source files the text came from.

[0078] A high level flow chart of the compression and shuffling module is shown at Fig. 9. At 901, object module II is read in. At 902 the text partition is processed. At 903 the other sections are processed. At 904 the header is updated. At 905, the object module is output.

[0079] Fig. 10 expands box 902. At 1001, the reference table, i.e. relocation information is gathered. At 1002, the branch targets are collected, because these are not to be compressed. At 1003, the software checks to see if there are more files in the source partition. If so, at 1004, the portion corresponding to the next file is retrieved. Then, at 1005, that portion is compressed. At 1006, file information in the source partition is updated. At 1007, the local symbol table is updated.

[0080] Once there are no more files in the source partition, the global symbol table is updated at 1008. Then, at 1009, address references in the text section are updated. Then at 1010, 256-bit shuffling is effected. Motivation for such shuffling will be discussed below.

[0081] Fig. 11 expands box 1005. First, it is determined at 1101 whether there are more instructions to be compressed. If so, a next instruction is retrieved at 1102. Subsequently each operation in the instruction is compressed at 1103 as per the tables in Figs. 5a and 5b and a scatter table is updated at 1108. The scatter table is a new data structure, required as a result of compression and shuffling, which will be explained further below. Then, at 1104, all of the operations in an instruction and the format bits of a subsequent instruction are combined as per Figs. 4a - 4e. Subsequently the relocation information in the reference table must be updated at 1105, if the current instruction contains an address. At 1106, information needed to update address references in the text section is gathered. At 1107, the compressed instruction is appended at the end of the output bit string and control is returned to box 1101. When there are no more instructions, control returns to box 1006.

[0082] Functions for handling compression are implemented in the various modules as listed below:

TABLE IV

	Name of module	identification of function performed
45	scheme_table	readable version of table of Figs. 5a and 5b
	comp_shuffle.c	256-bit shuffle, see box 1010
	comp_scheme.c	boxes 1103-1104
50	comp_bitstring.c	boxes 1005 & 1009
	comp_main.c	controls main flow of Figs. 9 and 10
55	comp_src.c, comp_reference.c, comp_misc.c, comp_btarget.c	miscellaneous support routines for performing other functions listed in Fig. 11

[0083] The scatter table, which is required as a result of the compression and shuffling of the invention, can be explained as follows.

[0084] The reference table contains a list of locations of addresses used by the instruction stream and corresponding list of the actual addresses listed at those locations. When the code is compressed, and when it is loaded, those addresses must be updated. Accordingly, the reference table is used at these times to allow the updating.

[0085] However, when the code is compressed and shuffled, the actual bits of the addresses are separated from each other and reordered. Therefore, the scatter table lists, for each address in the reference table, where EACH BIT is located. In the preferred embodiment the table lists, a width of a bit field, an offset from the corresponding index of the address in the source text, a corresponding offset from the corresponding index in the address in the destination text.

[0086] When object module III is loaded to run on the processor, the scatter table allows the addresses listed in the reference table to be updated even before the bits are deshuffled.

[0087] The scatter table contains, by way of example, as a set of scatter descriptors. Each scatter descriptor contains a set of triples (destination offset, width, source offset) and an unsigned integer indicating the number of triples in the descriptor.

[0088] For example, let us say we have a scatter descriptor with three triples (0,7,3), (7,4,15), (11, 5, 23). Let us say the source field is at position 320 in the bitfield of the text section. To get the bits of the actual address field, we do the following: bits 0 through 6 (7 bits) of the address field are from positions 323 (320+3) through 329 from the bitstring, bits 7 through 10 of the address field are from bits 335 through 338 of the bitstring, and bits 11 through 15 of the address field are from bits 343 through 347 of the bitstring. Thus the address field has length 16.

[0089] In the object module a list of reference descriptors is associated with a bitstring. Each reference descriptor refers to a bitfield in the bitstring. Each reference descriptor contains an index into the scatter table where we can find the scatter descriptor that has information about the way the bits of the bitfield are scattered in the bitstring. For example, if a bitfield has position 11 in a bitstring and the scatter descriptor corresponding to the bitfield has a single entry (0,18,0), then the actual source offset is obtained by adding the position and the source offset together: 11+0.

DECOMPRESSING THE INSTRUCTIONS

[0090] In order for the VLIW processor to process the instructions compressed as described above, the instructions must be decompressed. After decompression, the instructions will fill the instruction register, which has N issue slots,

N being 5 in the case of the preferred embodiment. Fig. 12 is a schematic of the decompression process. Instructions come from memory 1201, i.e. either from the main memory 104 or the instruction cache 105. The instructions must then be deshuffled 1201, which will be explained further below, before being decompressed 1203. After decompression 1203, the instructions can proceed to the CPU 1204.

[0091] Each decompressed operation has 2 format bits plus a 42 bit operation. The 2 format bits indicate one of the four possible operation lengths (unused issue slot, 26-bit, 34-bit, or 42-bit). These format bits have the same values as "Format" in section 5 above. If an operation has a size of 26 or 34 bits, the upper 8 or 16 bits are undefined. If an issue slot is unused, as indicated by the format bits, then all operation bits are undefined and the CPU has to replace the op code by a NOP op code (or otherwise indicate NOP to functional units).

[0092] Formally the decompressed instruction format is

<decompressed instruction> ::= { <decompressed operation> }N

<decompressed operation> ::= <operation:42> <format:2>

[0093] Operations have the format as in Table III (above).

[0094] Appendix A is VERILOG code which specifies the functioning of the decompression unit. VERILOG code is a standard format used as input to the VERILOG simulator produced by Cadence Design Systems, Inc. of San Jose, California. The code can also be input directly to the design compiler made by Synopsys of Mountain View California to create circuit diagrams of a decompression unit which will decompress the code. The VERILOG code specifies a list of pins of the decompression unit these are

TABLE V

# of pins in group	name of group of pins	description of group of pins
512	data512	512 bit input data word from memory, i.e. either from the instruction cache or the main memory

EP 0 843 848 B1

TABLE V (continued)

# of pins in group	name of group of pins	description of group of pins
5	32	input program counter
	44	output contents of issue slot 4
	44	output contents of issue slot 3
	44	output contents of issue slot 2
10	44	output contents of issue slot 1
	44	output contents of issue slot 0
	10	output duplicate of format bits in operations
	32	output first 32 bits pointed to by program counter
15	1	is it a branch target or not?
	1, each	input global pipeline control signals

20 [0095] Data512 is a double word which contains an instruction which is currently of interest. In the above, the program counter, PC is used to determine data512 according to the following algorithm:

```

A := {PC[31:8], 8'b0}
25      if PC[5] = 0 then

30          data512' := {M(A), M(A+32)}^
else data512' := {M(A+32), M(A)}

```

where

35 A is the address of a single word in memory which contains an instruction of interest;
8'b0 means 8 bits which are zeroed out
M(A) is a word of memory addressed by A;
M(A+32) is word of memory addressed by A+32;
40 data512' is the shuffled version of data 512

This means that words are swapped if an odd word is addressed.

[0096] Operations are delivered by the decompression unit in a form which is only partially decompressed, because the operation fields are not always in the same bit position. Some further processing has to be done to extract the operation fields from their bit position, most of which can be done best in the instruction decode stage of the CPU pipeline. For every operation field this is explained as follows:

- | | |
|-------|--|
| src1 | The src1 field is in a fixed position and can be passed directly to the register file as an address. Only the 32-bit immediate operation does not use the src1 field. In this case the CPU control will not use the src1 operand from the register file. |
| src2 | The src2 field is in a fixed position if it is used and can be passed directly to the register file as address. If it is not used it has an undefined value. The CPU control makes sure that a "dummy" src2 value read from the register file is not used. |
| guard | The guard field is in a fixed position if it is used and can be passed directly to the register file as an address. Simultaneously with register file access, the CPU control inspects the op code and format bits of the operation. If the operation is unguarded, the guard value read from the RF (reg- |

ister file) is replaced by the constant TRUE.

op code	Short or long op code and format bits are available in a fixed position in the operation. They are in bit position 21-30 plus the 2 format bits. They can be fed directly to the op code decode with maximum time for decoding.
5 dst	The dst field is needed very quickly in case of a 32-bit immediate operation with latency 0. This special case is detected quickly by the CPU control by inspecting bit 33 and the formal bits. In all other cases there is a full clock cycle available in the instruction decode pipeline state to decode where the dst field is in the operation (it can be in many places) and extract it.
10 32-bit immediate	If there is a 32-bit immediate it is in a fixed position in the operation. The 7 least significant bits are in the src2 field in the same location as a 7-bit parameter would be.
15 7-bit parameter	If there is a 7-bit parameter it is in the src2 field of the operation. There is one exception: the store with offset operation. For this operation, the 7-bit parameter can be in various locations and is multiplexed onto a special 7-bit immediate bus to the data cache.

BIT SWIZZLING

20 [0097] Where instructions are long, e.g. 512 bit double words, cache structure becomes complex. It is advantageous to swizzle the bits of the instructions in order to simplify the layout of the chip. Herein, the words swizzle and shuffle are used to mean the same thing. The following is an algorithm for swizzling bits.

25 **for (k=0; k<4; k=k+1)**
 for (i=0; i<8; i=i+1)
 for (j=0; j<8; j=j+1)
 begin
 word_shuffled[k*64+j*8+i] =
 word_unshuffled[(4*i+k)*8 + j];
 end

35
 40 where i, j, and k are integer indices; word_shuffled is a matrix for storing bits of a shuffled word; and word_unshuffled is matrix for storing bits of an unshuffled word.

CACHE STRUCTURE

45 [0098] Fig. 6a shows the functioning on input of a cache structure which is useful in efficient processing of VLIW instructions. This cache includes 16 banks 601-616 of 2k bytes each. These banks share an input bus 617. The caches are divided into two stacks. The stack on the left will be referred to as "low" and the stack on the right will be referred to as "high".
 50 [0099] The cache can take input in only one bank at a time and then only 4 bytes at a time. Addressing determines which 4 bytes of which bank are being filled. For each 512 bit word to be stored in the cache, 4 bytes are stored in each bank. A shaded portion of each bank is illustrated indicating corresponding portions of each bank for loading of a given word. These shaded portions are for illustration only. Any given word can be loaded into any set of corresponding portions of the banks.
 55 [0100] After swizzling according to the algorithm indicated above, sequential 4 byte portions of the swizzled word are loaded into the banks in the following order 608, 616, 606, 614, 604, 612, 602, 610, 607, 615, 605, 613, 603, 611, 601, 609. The order of loading of the 4 byte sections of the swizzled word is indicated by roman numerals in the boxes representing the banks.

[0101] Fig 6b shows how the swizzled word is read out from the cache. Fig. 6b shows only the shaded portions of the banks of the low stack. The high portion is analogous. Each shaded portion 601a-608a has 32 bits. The bits are loaded onto the output bus, called bus256low, using the connections shown, i.e. in the following order: 608a - bit0, 607a - bit 0, ..., 601a - bit 0; 608a - bit 1, 607a - bit1, ..., 601a -- bit 1; ...; 608a - bit 31, 607a - bit 31, ..., 601a - bit 31.

5 Using these connections, the word is automatically de-swizzled back to its proper bit order.

[0102] The bundles of wires, 620, 621, ..., 622 together form the output bus256 low. These wires pass through the cache to the output without crossing

[0103] On output, the cache looks like Fig. 7. The bits are read out from stack low 701 and stack high 702 under control of control unit 704 through a shift network 703 which assures that the bits are in the output order specified above. In this way the entire output of the 512 bit word is assured without bundles 620, 621, ... 622 and analogous wires crossing.

[0104] In the preceding a VLIW processor has been described that uses compressed instructions. The VLIW processor has an instruction issue register comprising a plurality of issue slots, each issue slot being for storing a respective operation, all of the operations starting execution in a same clock cycle. The VLIW processor has a plurality of functional units for executing the operations stored in the instruction register. The VLIW processor has a decompression unit for providing decompressed instructions to the instruction issue register, the decompression unit taking compressed instructions from a compressed instruction storage medium and decompressing the compressed instructions. At least one of the compressed instructions includes at least one operation, each operation being compressed according to a compression scheme which assigns a compressed operation length to that operation. The compressed operation length is chosen from a plurality of finite lengths, which finite lengths include at least two non-zero lengths, which of the finite lengths is chosen being dependent upon at least one feature of the operation.

[0105] Preferably, the set of operation lengths is {0, 26, 34, 42}. Also preferably the at least one feature is at least one of the following:

- 25 - abbreviated op code;
 - guarded or unguarded;
 - resultless;
 - immediate parameter with fixed number of bits; and
 - zeroary, unary, or binary.

30 The fixed number is preferably one of 7 and 32. Preferably the processor comprises a plurality of such instructions, of which one instruction is a branch target, which one instruction is not compressed. Preferably, each operation field within each instruction includes a sub-field specifying at least one of the following: a register file address of a first operand; a register file address of a second operand; a register file address of guard information; a register file address of a result; an immediate parameter; and an op code. Preferably, each instruction comprises a format field for specifying a plurality of respective formats, one respective format for each operation of a succeeding instruction. Preferably the compressed format comprises a format field specifying issue slots of the VLIW processor to be used by some instruction. Preferably at least one field specifies the operation. The field specifying the operation comprises at least one byte aligned sub-field. Preferably at least one operation part sub-field is located in a same byte with the format field. Thus 35 instructions may be aligned with a byte boundary, not just word boundaries. Preferably the format field may specify that more than a threshold quantity of issue slots are to be used and further comprises at least one first operation part sub-field located in a same byte with the format field, a plurality of sub-fields specifying operations, and at least one second operation part sub-field located in a byte separate from the other sub-fields.

40

45

50

55

APPENDIX A

```

5      // Verilog HDL for Icache. ic_decompression _behavioral
6
7      'define FIXED_FORMAT 10'b1010101010
8      'define NOP_OPERATION 42'b0
9
10     module    ic_decompression (data512, pc,
11                                operation4, operation1, operation2, operation1,
12                                operation0, format_out, first_word,
13                                format_ctrl0,
14                                reissue1, stall_in, freeze, reset, clk);
15
16     input [511:0] data512;
17     input [31:0] pc;
18     output [43:0] operation0, operation1, operation2, operation3, operation4;
19     reg [43:0] operation0, operation1, operation2, operation3, operation4;
20     output [31:0] first_word;
21     reg [31:0] first_word;
22     input format_ctrl0;
23     input reissue1, freeze, reset, clk;
24     wire [9:0] format_out1;
25     output [9:0] format_out;
26     reg [9:0] format_outA, format_B;
27     input stall_in;
28
29     // local
30     reg [9:0] format_out0;
31     reg [31:0] pc_p;
32     reg format_ctrl1;
33
34     reg [9:0] format;
35     reg used0, used1, used2, used3, used4;
36     reg [1:0] size0, size1, size2, size3, size4;
37
38     reg [511:0] data512shift;
39     reg [255:0] data256;
40     reg [2:0] pos1, pos2, pos3, pos4, pos_ext;
41
42     reg [25:0] fix0, fix1, fix2, fix3, fix4;
43     reg [79:0] extension0, extension1, extension2, extension3, extension4;
44     reg [15:0] ext0, ext1, ext2, ext3, ext4;
45     reg reset_p;
46
47     // format pipe
48     always @(posedge clk) . . .
49     begin
50       reset_p <= reset;
51
52       if (reset_p)
53         begin
54           // force NOP operations on instructions
55           operation0[42] <= 'ONE;
56           operation0[43] <= 'ONE;
57           operation1[42] <= 'ONE;
58           operation1[43] <= 'ONE;
59           operation2[42] <= 'ONE;
60           operation2[43] <= 'CNE;
61           operation3[42] <= 'CNE;
62           operation3[43] <= 'ONE;
63           operation4[42] <= 'ONE;
64           operation4[43] <= 'CNE;
65         end
66       else if (!stall_in)
67         begin
68
69         pc_p <= pc;
70         format_ctrl1 <= format_ctrl0;
71
72       end
73     end
74   endmodule

```

```

if (format_ctrl)
    format = 'FIXED_FORMAT;
else
    format = format_out;

5      used0 = -(format[1] & format[0]);
size0 = used0 ? format[1:0] : 2'b0;
        used1 = -(format[3] & format[2]);
size1 = used1 ? format[3:2] : 2'b0;
10     used2 = -(format[5] & format[4]);
size2 = used2 ? format[5:4] : 2'b0;
        used3 = -(format[7] & format[6]);
size3 = used3 ? format[7:6] : 2'b0;
15     used4 = -(format[9] & format[8]);
size4 = used4 ? format[9:8] : 2'b0;

// first alignment stage
// rotate the 512 bit word right over a distance between 0 and 64 byte
20     //
// the rotate is implemented here by swapping the left and right word if the
// distance is more than 32 byte and then perform a right shift over a distance
// between
// 0 and 32 byte.

25     data512shift = pc_p[5] ?
                (data512[255:0], data512[511:256]) >> (pc_p[4:0], 3'b0) :
                data512 >> (pc_p[4:0], 3'b0);

data256 = data512shift[255:0];

// extract format bits
30     format_out0 = data256[9:0];

// access first word
first_word <= data256[31:0];

// Notes: - the value for pos_ext==0 is don't care
//         - for values pos_ext < 5 , less than 80 bits are needed
35     //

// determine the position of issue slots

//pos0 = 0;
pos1 = used0;
pos2 = used0 + used1;
40     pos3 = used0 + used1 + used2;
pos4 = used0 + used1 + used2 + used3;

// mux the fixed part of issue slots, combine the 24-bit part and the 2-bit part
fix0 = used0 ? (data256[15:14], data256[(0+1)*24+15 : 0*24+16]) : 'NOP_OPERATION;
45     ;

fix1 = used1 ?
        (
            pos1 == 0 ? (data256[15:14], data256[(0+1)*24+15 : 0*24+16]) :
            (data256[13:12], data256[(1+1)*24+15 : 1*24+16])
        )
50     : 'NOP_OPERATION;

fix2 = used2 ?
        (

```

```

    pos2 == 0 ? {data256[15:14], data256[(0+1)*24+15 : 0*24+16]} :
    pos2 == 1 ? {data256[13:12], data256[(1+1)*24+15 : 1*24+16]} :
    (data256[11:10], data256[(2+1)*24+15 : 2*24+16]}
)
: 'NOP_OPERATION;

fix3 = used3 ?
(
    pos3 == 0 ? {data256[15:14], data256[(0+1)*24+15 : 0*24+16]} :
    pos3 == 1 ? {data256[13:12], data256[(1+1)*24+15 : 1*24+16]} :
    pos3 == 2 ? {data256[11:10], data256[(2+1)*24+15 : 2*24+16]} :
    (data256[95:94], data256[(0+1)*24+95 : 0*24+96]}
)
: 'NOP_OPERATION;

fix4 = used4 ?
(
    pos4 == 0 ? {data256[15:14], data256[(0+1)*24+15 : 0*24+16]} :
    pos4 == 1 ? {data256[13:12], data256[(1+1)*24+15 : 1*24+16]} :
    pos4 == 2 ? {data256[11:10], data256[(2+1)*24+15 : 2*24+16]} :
    pos4 == 3 ? {data256[95:94], data256[(0+1)*24+95 : 0*24+96]} :
    (data256[93:92], data256[(1+1)*24+95 : 1*24+96]}
)
: 'NOP_OPERATION;

// determine the position of the extension part
pos_ext = used0 + used1 + used2 + used3 + used4;

// determine the extension
extension0 =
(
    pos_ext == 0 ? data256[0*24+80-1+16 : 0*24+16] :
    pos_ext == 1 ? data256[1*24+80-1+16 : 1*24+16] :
    pos_ext == 2 ? data256[2*24+80-1+16 : 2*24+16] :
    pos_ext == 3 ? data256[3*24+80-1+16 : 3*24+16] :
    pos_ext == 4 ? data256[1*24+80-1+96 : 1*24+96] :
    data256[2*24+80-1+96 : 2*24+96];
)

// shift the Extension part
extension1 = extension0 >> (size0 . 3'b0);
extension2 = extension1 >> (size1 . 3'b0);
extension3 = extension2 >> (size2 . 3'b0);
extension4 = extension3 >> (size3 . 3'b0);

ext0 = extension0[15:0];
ext1 = extension1[15:0];
ext2 = extension2[15:0];
ext3 = extension3[15:0];
ext4 = extension4[15:0];

// assemble instruction
//operation0 <= (format_out0[1:0], ext0, fix0);
//operation1 <= (format_out0[3:2], ext1, fix1);
//operation2 <= (format_out0[5:4], ext2, fix2);
//operation3 <= (format_out0[7:6], ext3, fix3);
//operation4 <= (format_out0[9:8], ext4, fix4);
operation0 <= (format[1:0], ext0, fix0);
operation1 <= (format[3:2], ext1, fix1);
operation2 <= (format[5:4], ext2, fix2);
operation3 <= (format[7:6], ext3, fix3);
operation4 <= (format[9:8], ext4, fix4);

if (-freeze | reissue1)
begin
    format_out1 <= format_out0;
end

```

```

5      if (-freeze)
begin
6          format_p <= format_outA;
end

7      end
8      end

9      assign format_out = reissue1 ? format_p : format_outA;
10     assign format_out1 = format;

11     endmodule
12
13
14
15

```

Claims

20

1. A VLIW processor for using compressed instructions, the processor comprising

25

- an instruction issue register (154) comprising a plurality of issue slots, each issue slot being for storing a respective operation, all of the operations starting execution in a same clock cycle;
- a plurality of functional units (151, 152, 153) for executing the operations stored in the instruction issue register (154);
- a decompression unit (155) for providing a decompressed instruction to the instruction issue register (154), the decompression unit (155) taking a compressed instruction from a compressed instruction storage medium (103) and decompressing the compressed instruction, the compressed instruction including at least two operations, each of said at least two operations being compressed to a respective compressed operation length,

30

characterized in that the decompression unit (155) is arranged to decompress operations with respective compressed operation lengths chosen from a plurality of finite lengths, which finite lengths include at least two non-zero lengths.

35

2. The processor of claim 1, the decompression unit (155) being arranged to take a format field from the compressed instruction storage medium (103), the format field specifying the respective compressed operation length for each operation of the compressed instruction, the decompression unit (155) decompressing the operations of the compressed instruction according to the format field.

40

3. The processor of 2 wherein the format field has N sub-fields, N being the number of issue slots, each sub-field specifying a compressed operation length for a respective issue slot, **characterized in that** the sub-fields each contain at least two bits.

45

4. The processor of Claim 2 or 3, the decompression unit (155) being arranged for

50

- reading a preceding compressed instruction from the compressed instruction storage medium (103) together with the format field.
- starting decompression of the preceding compressed instruction and subsequently
- taking the compressed instruction from the compressed instruction memory and starting decompression of the compressed instruction according to the format field taken from the compressed instruction medium (103) together with the preceding compressed instruction.

55

5. The processor of claim 2, 3 or 4, the decompression unit (155) taking the format field from the compressed instruction storage medium (103) in a memory access unit, the memory access unit also comprising at least one operation part sub-field, the decompression unit (155) integrating the operation part sub-field in at least one of the operations of the decompressed instruction.

6. The VLIW processor of claim 1, wherein: the decompression unit (155) is arranged for taking a stream of compressed instructions from the compressed instruction storage medium (103) and decompressing the compressed instructions, the stream of instructions comprising: a first instruction including a format field which specifies an instruction compression format, and wherein the stream of instructions comprises a second instruction, taken from the compressed instruction storage medium (103) following the first instruction, the decompression unit (155) being arranged to decompress the second instruction according to the format field in the first instruction.
- 5
7. A method of producing compressed code for running on a VLIW processor, the method comprising the steps of
- 10
- receiving an instruction comprising at least two operations
 - compressing each of said at least two operations of the instruction according to a respective compression scheme which assigns a respective compressed operation length to the relevant operation,
- 15
- characterized in that** the compressed operation length is chosen from a plurality of finite lengths, which finite lengths include at least two non-zero lengths, which of the finite lengths is chosen depending upon at least one feature of the operation.
8. The method of claim 7 applied to a stream of instructions including said instruction, the method comprising the step of determining for each instruction whether that instruction is a branch target (i1) of a branch from another instruction of the stream of instructions, and compressing only those instructions which are not branch targets.
- 20
9. The method of Claim 7 or 8, comprising the step of producing a format field, the format field specifying a respective format for each operation of the instruction according to the compressed operation length chosen for that operation.
- 25
10. The method of 9 wherein the format field has N sub-fields, N being the number of issue slots, each sub-field specifying a compressed operation length for a respective issue slot, **characterized in that** the sub-fields each contain at least two bits.
- 30
11. The method of Claim 7, 8, 9 or 10, comprising the step of storing a compressed instruction containing the compressed operations in a computer readable compressed instruction storage medium (103).
12. The method of claim 11, when dependent on claim 9, comprising compressing a further instruction, for execution preceding execution of the instruction, the format field for the instruction being stored for fetching with the further instruction.
- 35
13. The method of claim 11, when dependent on claim 9, or claim 12 the compressed instruction storage medium (103) having memory access units, the format field being stored in a same memory access unit with at least one operation part sub-field of at least one of the operations of the instruction.
- 40
14. The method of claim 11 when dependent on claim 9, wherein the receiving step comprises
- 45

 - receiving a stream of instructions each comprising a plurality of operations wherein the stream comprises a first instruction for execution preceding a second instruction from the stream, the format field corresponding to the second instruction and the compressed instruction corresponding to the first instruction being stored in the compressed instruction storage medium (103) for combined fetching during execution of the stream, prior to fetching of the compressed instruction corresponding to the second instruction.
15. The method of claim 14 comprising the step of determining for each instruction whether that instruction is a branch target (i2) of a branch from another instruction of the stream of instructions, and compressing only those instructions which are not branch targets.
- 50
16. A computer programmed to execute the method according to any one of the claims 7 to 15.

55 **Revendications**

1. Processeur VLIW pour utiliser des instructions comprimées, le processeur comprenant

- un registre d'instructions émises (154) comprenant une pluralité de champs d'instructions émises, chaque champ d'instruction émise servant à stocker une opération respective, toutes les opérations commençant leur exécution dans un même cycle d'horloge ;
- une pluralité d'unités fonctionnelles (151, 152, 153) pour exécuter les opérations stockées dans le registre d'instructions émises (154) ;
- une unité de décompression (155) pour fournir une instruction décomprimée au registre d'instructions émises (154), l'unité de décompression (155) extrayant une instruction comprimée d'un support de stockage d'instructions comprimées (103) et décompressant l'instruction comprimée, l'instruction comprimée comprenant au moins deux opérations, chacune desdites au moins deux opérations étant comprimée à une longueur d'opération comprimée respective ;

caractérisé en ce que l'unité de décompression (155) est à même de décompresser des opérations avec des longueurs d'opération comprimée respectives choisies parmi une pluralité de longueurs finies, lesquelles longueurs finies incluent au moins deux longueurs non zéro.

2. Processeur suivant la revendication 1, l'unité de décompression (155) étant à même d'extraire un champ de format du support de stockage d'instructions comprimées (103), le champ de format spécifiant la longueur d'opération comprimée respective pour chaque opération de l'instruction comprimée, l'unité de décompression (155) décompressant les opérations de l'instruction comprimée en fonction du champ de format.
3. Processeur suivant la revendication 2 dans lequel le champ de format comprend N sous-champs, N étant le nombre de champs d'instructions émises, chaque sous-champ spécifiant une longueur d'opération comprimée pour un champ d'instruction émise respectif, **caractérisé en ce que** les sous-champs contiennent chacun au moins deux bits.
4. Processeur suivant la revendication 2 ou 3, l'unité de décompression (155) étant à même de
 - d'extraire une instruction comprimée précédente du support de stockage d'instructions comprimées (103) ainsi que le champ de format ;
 - de commencer la décompression de l'instruction comprimée précédente, et ensuite
 - d'extraire l'instruction comprimée de la mémoire d'instructions comprimées et de commencer la décompression de l'instruction comprimée suivant le champ de format extrait du support de stockage d'instructions comprimées (103) ainsi que l'instruction comprimée précédente.
5. Processeur suivant la revendication 2, 3 ou 4, l'unité de décompression (155) extrayant le champ de format du support de stockage d'instructions comprimées (103) dans une unité d'accès à la mémoire, l'unité d'accès à la mémoire comprenant également au moins un sous-champ de partie d'opération, l'unité de décompression (155) intégrant le sous-champ de partie d'opération dans au moins une des opérations de l'instruction décomprimée.
6. Processeur VLIW suivant la revendication 1, dans lequel
 - l'unité de décompression (155) est à même d'extraire un train d'instructions comprimées du support de stockage d'instructions comprimées (103) et de décompresser les instructions comprimées, le train d'instructions comprenant : une première instruction incluant un champ de format qui spécifie un format de compression d'instructions, et dans lequel
 - le train d'instructions comprend une deuxième instruction, extraite du support de stockage d'instructions comprimées (103) suivant la première instruction, l'unité de décompression (155) étant à même de décompresser la deuxième instruction suivant le champ de format dans la première instruction.
7. Procédé de production de code comprimé à exécuter sur un processeur VLIW, le procédé comprenant les étapes suivantes :
 - la réception d'une instruction comprenant au moins deux opérations ;
 - la compression de chacune desdites au moins deux opérations de l'instruction suivant un schéma de compression respectif qui attribue une longueur d'opération comprimée respective à l'opération concernée,

caractérisé en ce que la longueur d'opération comprimée est choisie parmi une pluralité de longueurs finies, lesquelles longueurs finies incluent au moins deux longueurs non zéro, celle parmi les longueurs finies qui est choisie dépendant d'au moins une caractéristique de l'opération.

8. Procédé suivant la revendication 7 appliqué à un train d'instructions comprenant ladite instruction, le procédé comprenant l'étape de détermination pour chaque instruction de si cette instruction est une cible de branchement (i1) d'un branchement d'une autre instruction du train d'instructions, et de compression des seules instructions qui ne sont pas des cibles de branchement.
- 5 9. Procédé suivant la revendication 7 ou 8 comprenant l'étape de production d'un champ de format, le champ de format spécifiant un format respectif pour chaque opération de l'instruction suivant la longueur d'opération comprimée choisie pour cette opération.
- 10 10. Procédé suivant la revendication 9 dans lequel le champ de format comprend N sous-champs, N étant le nombre de champs d'instructions émises, chaque sous-champ spécifiant une longueur d'opération comprimée pour un champ d'instruction émise respectif, **caractérisé en ce que** les sous-champs contiennent chacun au moins deux bits.
- 15 11. Procédé suivant la revendication 7, 8, 9 ou 10 comprenant l'étape de stockage d'une instruction comprimée contenant les opérations comprimées dans un support de stockage d'instructions comprimées lisible par ordinateur (103).
- 20 12. Procédé suivant la revendication 11, lorsqu'elle dépend de la revendication 9, comprenant la compression d'une autre instruction, pour exécution avant l'exécution de l'instruction, le champ de format pour l'instruction étant stocké pour extraction avec l'autre instruction.
- 25 13. Procédé suivant la revendication 11, lorsqu'elle dépend de la revendication 9, ou la revendication 12, le support de stockage d'instructions comprimées (103) comprenant des unités d'accès à la mémoire, le champ de format étant stocké dans une même unité d'accès à la mémoire avec au moins un sous-champ de partie d'opération d'au moins une des opérations de l'instruction.
14. Procédé suivant la revendication 11, lorsqu'elle dépend de la revendication 9, dans lequel l'étape de réception comprend
- 30 - la réception d'un train d'instructions comprenant chacune une pluralité d'opérations ;
dans laquelle le train comprend une première instruction pour exécution avant une deuxième instruction du train, le champ de format correspondant à la deuxième instruction et l'instruction comprimée correspondant à la première instruction étant stockés dans le support de stockage d'instructions comprimées (103) pour une extraction combinée pendant l'exécution du train, avant l'extraction de l'instruction comprimée correspondant à la deuxième instruction.
- 35 15. Procédé suivant la revendication 14 comprenant l'étape de détermination pour chaque instruction de si cette instruction est une cible de branchement (i1) d'un branchement d'une autre instruction du train d'instructions, et de compression des seules instructions qui ne sont pas des cibles de branchement.
- 40 16. Ordinateur programmé pour exécuter le procédé suivant l'une quelconque des revendications 7 à 15.

45 Patentansprüche

1. VLIW-Prozessor für die Verwendung komprimierter Befehle, wobei der Prozessor Folgendes umfasst:
- 50 - ein Befehlsausgaberegister (154) mit einer Vielzahl von Issueslots, wobei jeder Issueslot zur Speicherung einer jeweiligen Operation dient und alle Operationen in dem selben Taktzyklus mit der Ausführung beginnen;
- eine Vielzahl von Funktionseinheiten (151, 152, 153) zur Ausführung der im Befehlsausgaberegister (154) gespeicherten Operationen;
- eine Dekompressionseinheit (155), um das Befehlsausgaberegister (154) mit einem dekomprimierten Befehl zu versorgen, wobei die Dekompressionseinheit (155) einem Speichermedium für komprimierte Befehle (103) einen komprimierten Befehl entnimmt und den komprimierten Befehl dekomprimiert, und wobei der komprimierte Befehl mindestens zwei Operationen enthält, wobei jede der mindestens zwei Operationen auf eine entsprechende komprimierte Operationslänge komprimiert ist,

dadurch gekennzeichnet, dass die Dekompressionseinheit (155) dafür eingerichtet ist, Operationen mit jeweils aus einer Vielzahl endlicher Längen gewählten komprimierten Operationslängen zu dekomprimieren, wobei die endlichen Längen mindestens zwei Längen enthalten, die nicht Null sind.

- 5 2. Prozessor nach Anspruch 1, wobei die Dekompressionseinheit (155) dafür eingerichtet ist, dem Speichermedium für komprimierte Befehle (103) ein Formatfeld zu entnehmen, wobei das Formatfeld die jeweilige komprimierte Operationslänge für jede Operation des komprimierten Befehls spezifiziert, und wobei die Dekompressionseinheit (155) die Operationen des komprimierten Befehls gemäß dem Formatfeld dekomprimiert.
- 10 3. Prozessor nach Anspruch 2, wobei das Formatfeld N Teifelder hat, wobei N die Anzahl der Issueslots ist und jedes Teifeld eine komprimierte Operationslänge für einen jeweiligen Issueslot spezifiziert, **dadurch gekennzeichnet, dass** die Teifelder jeweils mindestens zwei Bits enthalten.
- 15 4. Prozessor nach Anspruch 2 oder 3, wobei die Dekompressionseinheit (155) dafür eingerichtet ist,
 - dem Speichermedium für komprimierte Befehle (103) zusammen mit dem Formatfeld einen vorhergehenden komprimierten Befehl zu entnehmen,
 - mit dem Dekomprimieren des vorhergehenden komprimierten Befehls zu beginnen, und anschließend
 - dem Speicher für komprimierte Befehle den komprimierten Befehl zu entnehmen und mit dem Dekomprimieren des komprimierten Befehls gemäß dem Formatfeld zu beginnen, das dem Speichermedium für komprimierte Befehle (103) zusammen mit der vorhergehenden komprimierten Befehl entnommen wurde.
- 20 5. Prozessor nach Anspruch 2, 3 oder 4, wobei die Dekompressionseinheit (155) das Formatfeld dem Speichermedium für komprimierte Befehle (103) in einer Speicherzugriffseinheit entnimmt, wobei die Speicherzugriffseinheit weiterhin mindestens ein Operationsteifeld umfasst und die Dekompressionseinheit (155) das Operationsteifeld in mindestens eine der Operationen des dekomprimierten Befehls einbezieht.
- 25 6. VLIW-Prozessor nach Anspruch 1, wobei die Dekompressionseinheit (155) dafür eingerichtet ist, dem Speichermedium für komprimierte Befehle (103) einen Strom komprimierter Befehle zu entnehmen und die komprimierten Befehle zu dekomprimieren, und wobei der Befehlsstrom einen ersten Befehl umfasst, der ein Formatfeld enthält, welches ein Befehlskompressionsformat spezifiziert, wobei der Befehlsstrom einen zweiten Befehl umfasst, der dem Speichermedium für komprimierte Befehle (103) nach dem ersten Befehl entnommen wird, und wobei die Dekompressionseinheit (155) dafür eingerichtet ist, den zweiten Befehl gemäß dem Formatfeld in dem ersten Befehl zu dekomprimieren.
- 30 7. Verfahren zur Erzeugung eines komprimierten Codes, der auf einem VLIW-Prozessor abläuft, wobei das Verfahren folgende Schritte umfasst:
 - Empfangen eines Befehls mit mindestens zwei Operationen
 - Komprimieren jeder der genannten mindestens zwei Operationen des Befehls gemäß einem entsprechenden Kompressionsschema, das der betreffenden Operation eine entsprechende komprimierte Operationslänge zuordnet,
- 35 **dadurch gekennzeichnet, dass** die komprimierte Operationslänge aus einer Vielzahl endlicher Längen ausgewählt wird, wobei die endlichen Längen mindestens zwei Längen enthalten, die nicht Null sind, und die Auswahl einer der endlichen Längen von mindestens einem Merkmal der Operation abhängt.
- 40 8. Verfahren nach Anspruch 7, angewendet auf einen Befehlsstrom, der den genannten Befehl enthält, wobei das Verfahren den Schritt umfasst, für jeden Befehl zu bestimmen, ob dieser Befehl das Sprungziel (i1) einer Verzweigung von einem anderen Befehl des Befehlsstroms ist, und nur solche Befehle zu komprimieren, die keine Sprungziele sind.
- 45 9. Verfahren nach Anspruch 7 oder 8, umfassend den Schritt zur Erzeugung eines Formatfeldes, wobei das Formatfeld das jeweilige Format für jede Operation des Befehl gemäß der für diese Operation gewählten komprimierten Operationslänge spezifiziert.
- 50 10. Verfahren nach Anspruch 9, wobei das Formatfeld N Teifelder hat, N die Anzahl der Issueslots ist, und jedes Teifeld eine komprimierte Operationslänge für einen jeweiligen Issueslot spezifiziert, **dadurch gekennzeichnet,**

EP 0 843 848 B1

dass die Teilsteller jeweils mindestens zwei Bits enthalten.

5 **11.** Verfahren nach Anspruch 7, 8, 9 oder 10, umfassend den Schritt, einen die komprimierten Operationen enthaltenden komprimierten Befehl in einem computerlesbaren Speichermedium für komprimierte Befehle (103) zu speichern.

10 **12.** Verfahren nach Anspruch 11, wenn in Abhängigkeit von Anspruch 9, umfassend das Komprimieren eines weiteren Befehls, der vor dem Ausführen des vorangehenden Befehls auszuführen ist, wobei das Formatfeld für den Befehl zur Abholung mit dem weiteren Befehl gespeichert wird.

15 **13.** Verfahren nach Anspruch 11, wenn in Abhängigkeit von Anspruch 9, oder Anspruch 12, wobei das Speichermedium für komprimierte Befehle (103) Speicherzugriffseinheiten hat, und das Formatfeld in derselben Speicherzugriffseinheit mit mindestens einem Operationsteilfeld von mindestens einer der Operationen des Befehls gespeichert wird.

20 **14.** Verfahren nach Anspruch 11, wenn in Abhängigkeit von Anspruch 9, wobei der Schritt des Empfangens Folgendes umfasst:

- Empfangen eines Befehlsstroms mit einer Vielzahl von Operationen,

25 wobei der Strom einen ersten Befehl zur Ausführung vor einem zweiten Befehl aus dem Strom umfasst, das zum zweiten Befehl gehörende Formatfeld und der zum ersten Befehl gehörende komprimierte Befehl im Speichermedium für komprimierte Befehle (103) gespeichert werden, um während der Ausführung des Stroms zusammen abgerufen zu werden, bevor der zum zweiten Befehl gehörende komprimierte Befehl abgerufen wird.

30 **15.** Verfahren nach Anspruch 14, umfassend den Schritt, für jeden Befehl zu bestimmen, ob dieser Befehl das Sprungziel (i1) einer Verzweigung von einem anderen Befehl des Befehlsstroms ist, und nur solche Befehle zu komprimieren, die keine Sprungziele sind.

35 **16.** Computer, der für die Ausführung des Verfahrens gemäß einem der Ansprüche 7 bis 15 programmiert ist.

40

45

50

55

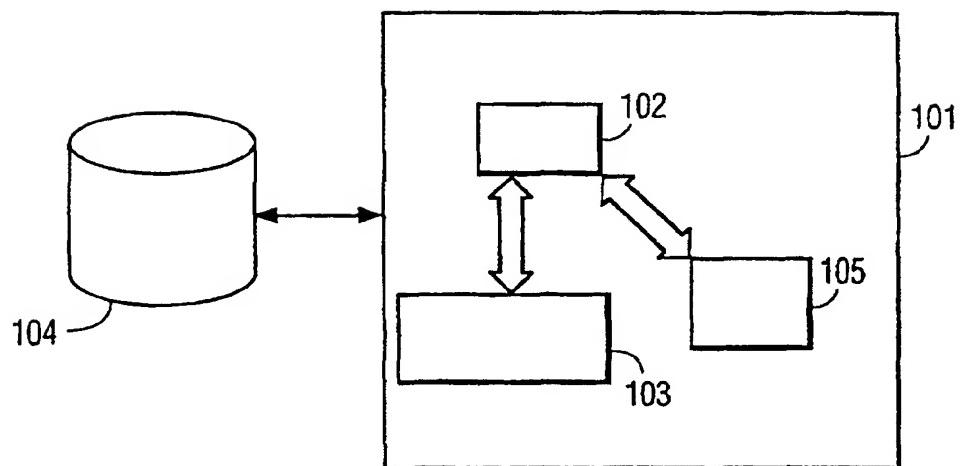


FIG. 1A

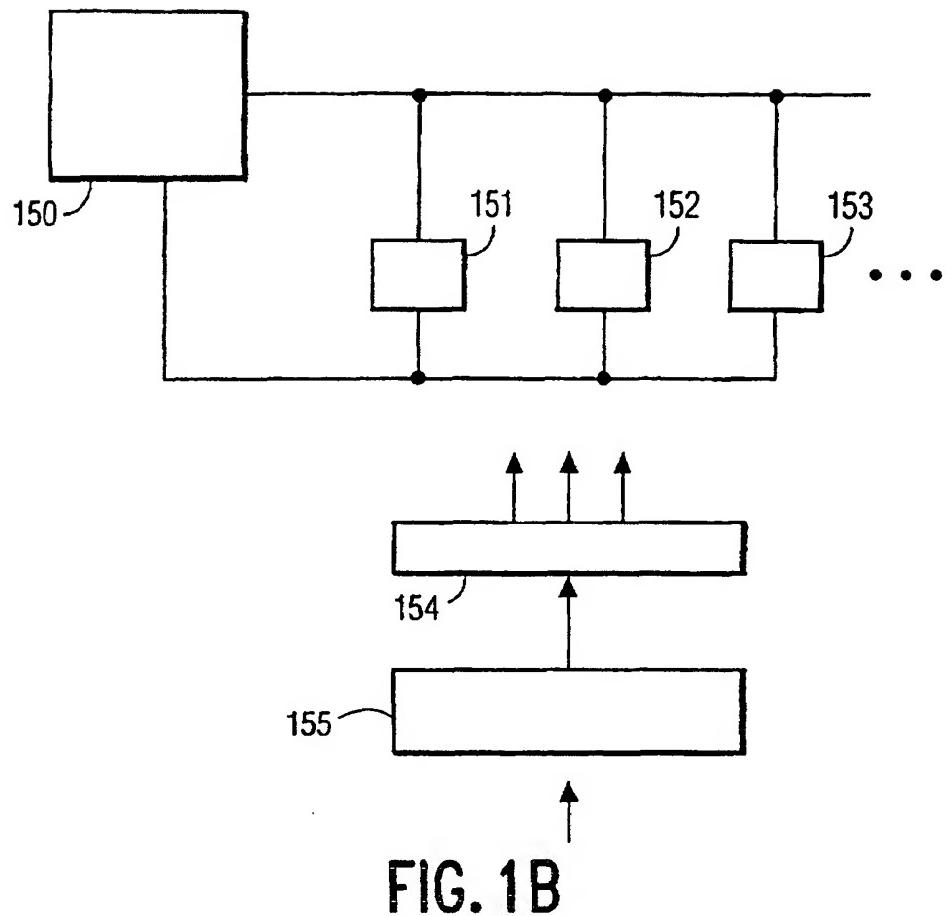


FIG. 1B



FIG. 2A



FIG. 2B



FIG. 2C



FIG. 2D



FIG. 2E

1- 2- 3- 4-

FIG. 3

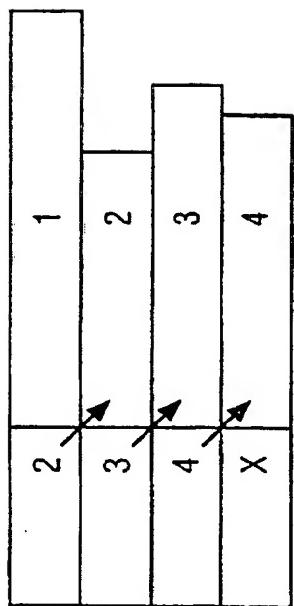


FIG. 4A

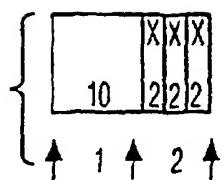


FIG. 4B

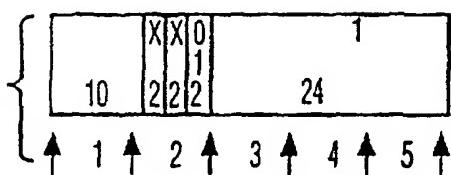


FIG. 4C

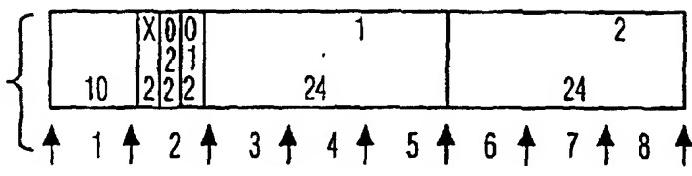


FIG. 4D

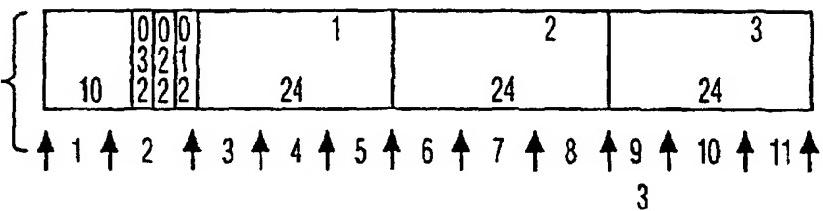


FIG. 4E

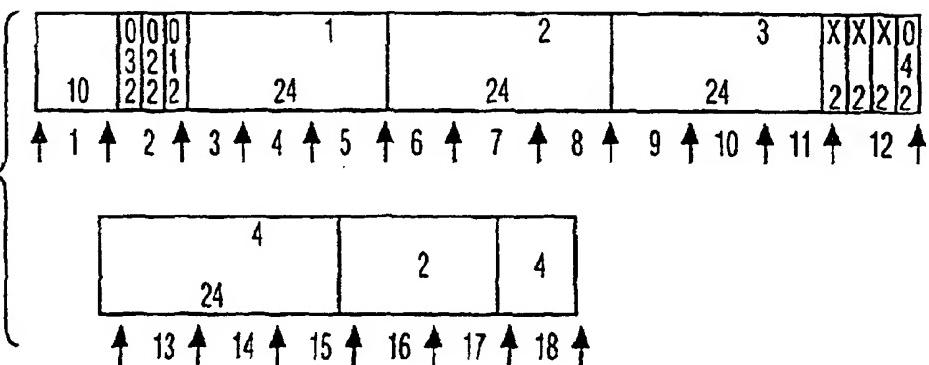
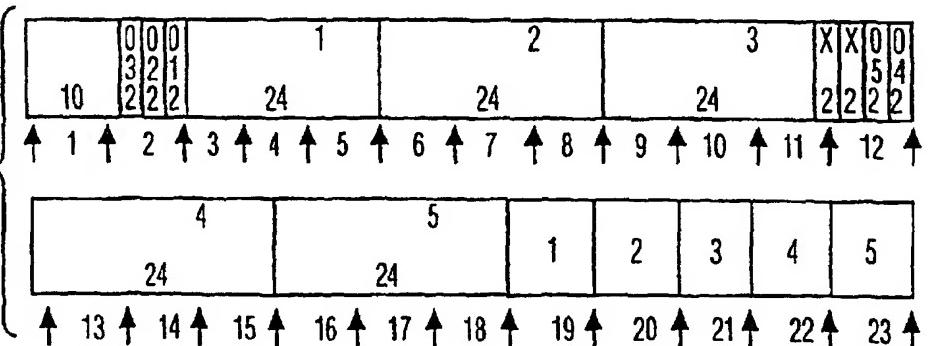


FIG. 4F



		24-bit	24-bit	2-bit	
26-		0-6	7-3	14-20	21-23
				24-25	26-34-41
	<binary-unguarded-short>	src1[0:6]	src2[0:6]	ds[0:6]	opcode[0:2]
	<unary-param7-unguarded-short>	src1[0:6]	param[0:6]	ds[0:6]	opcode[3:4]
	<binary-unguarded-param7-resultless-short>	src1[0:6]	src2[0:6]	param[0:6]	opcode[0:2]
	<unary-short>	src1[0:6]	src2[0:6]	param[0:6]	opcode[3:4]
34-				guard[0:6]	opcode[0:2]
	<binary-short> iadd, etc.	src1[0:6]	src2[0:6]	guard[0:6]	opcode[3:4]
	<unary-param7-short>	src1[0:6]	param[0:6]	guard[0:6]	opcode[0:2]
	<binary-param7-resultless-short>	src1[0:6]	src2[0:6]	guard[0:6]	opcode[3:4]
	<binary-unguarded>	src1[0:6]	src2[0:6]	ds[0:6]	opcode[0:2]
	<binary-resultless>	src1[0:6]	src2[0:6]	guard[0:6]	opcode[3:4]
	<unary-param7-un-guarded>	src1[0:6]	param[0:6]	ds[0:6]	opcode[0:2]
	<unary>	src1[0:6]	ds[0:6]	guard[0:6]	opcode[3:4]

FIG. 5A

42-						
<binary-param7-resultless>	src1[0:6]	src2[0:6]	guard[0:6]	opcode[0:2]	opcode[3:4]	opcode[5:7]SXX100param[0:6]
<binary>	src1[0:6]	src2[0:6]	guard[0:6]	opcode[0:2]	opcode[3:4]	opcode[5:7]XL0101 ds[0:6]
<unary-param7>	src1[0:6]	param[0:6]	guard[0:6]	opcode[0:2]	opcode[3:4]	opcode[5:7]SL1101 ds[0:6]
<zeroary-param32>	param[7:13]	param[0:6]	ds[0:6]	param[14:16]	param[17:18]	param[19:23]XX1 param[24:31]
<zeroary-param32-resultless>	param[7:13]	param[0:6]	guard[0:6]	param[14:16]	param[17:18]	param[19:23]000 param[24:31]
<zeroary-param32-resultless>	param[7:13]	param[0:6]	guard[0:6]	param[14:16]	param[17:18]	param[19:23]100 param[24:31]

FIG. 5B

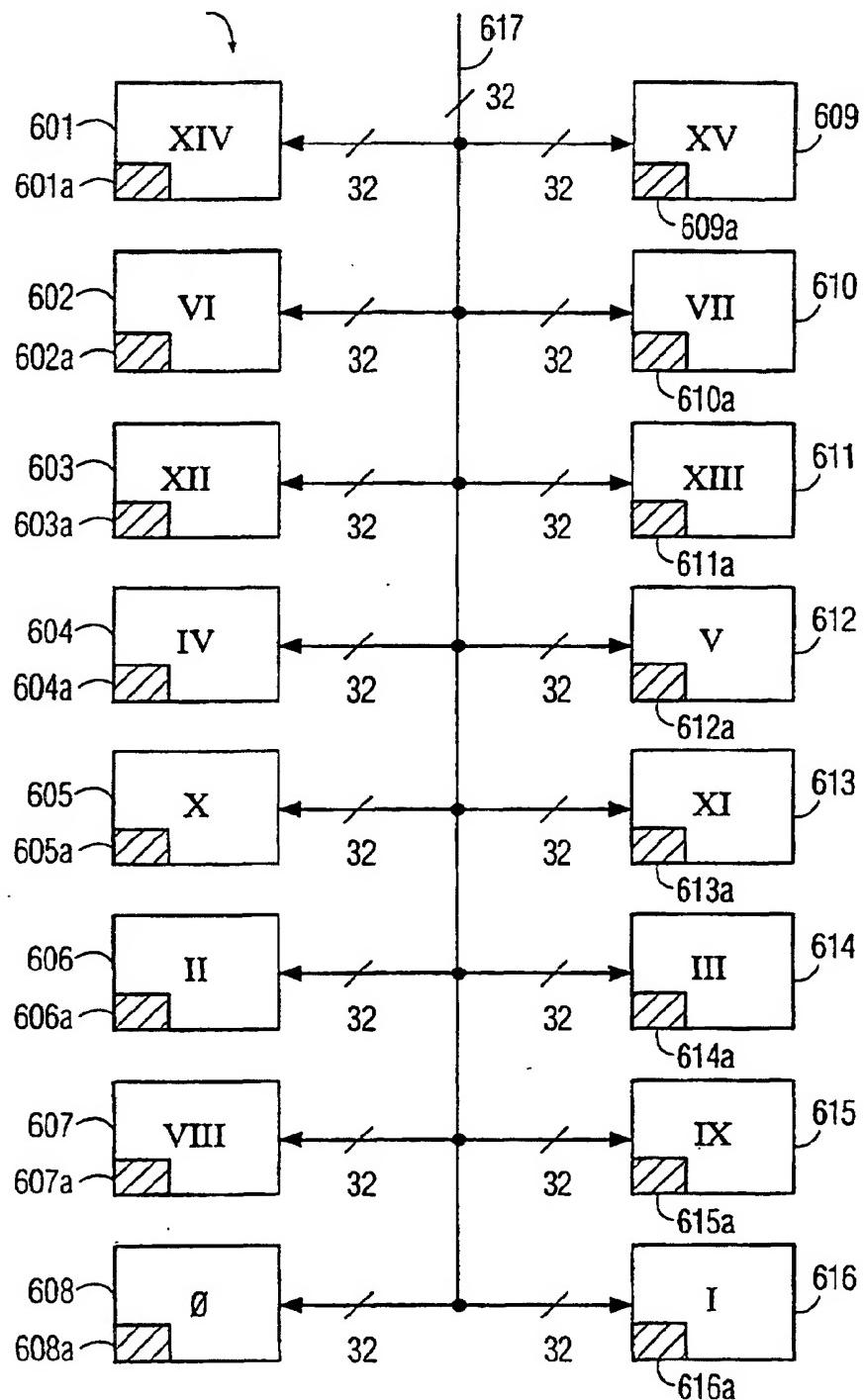


FIG. 6A

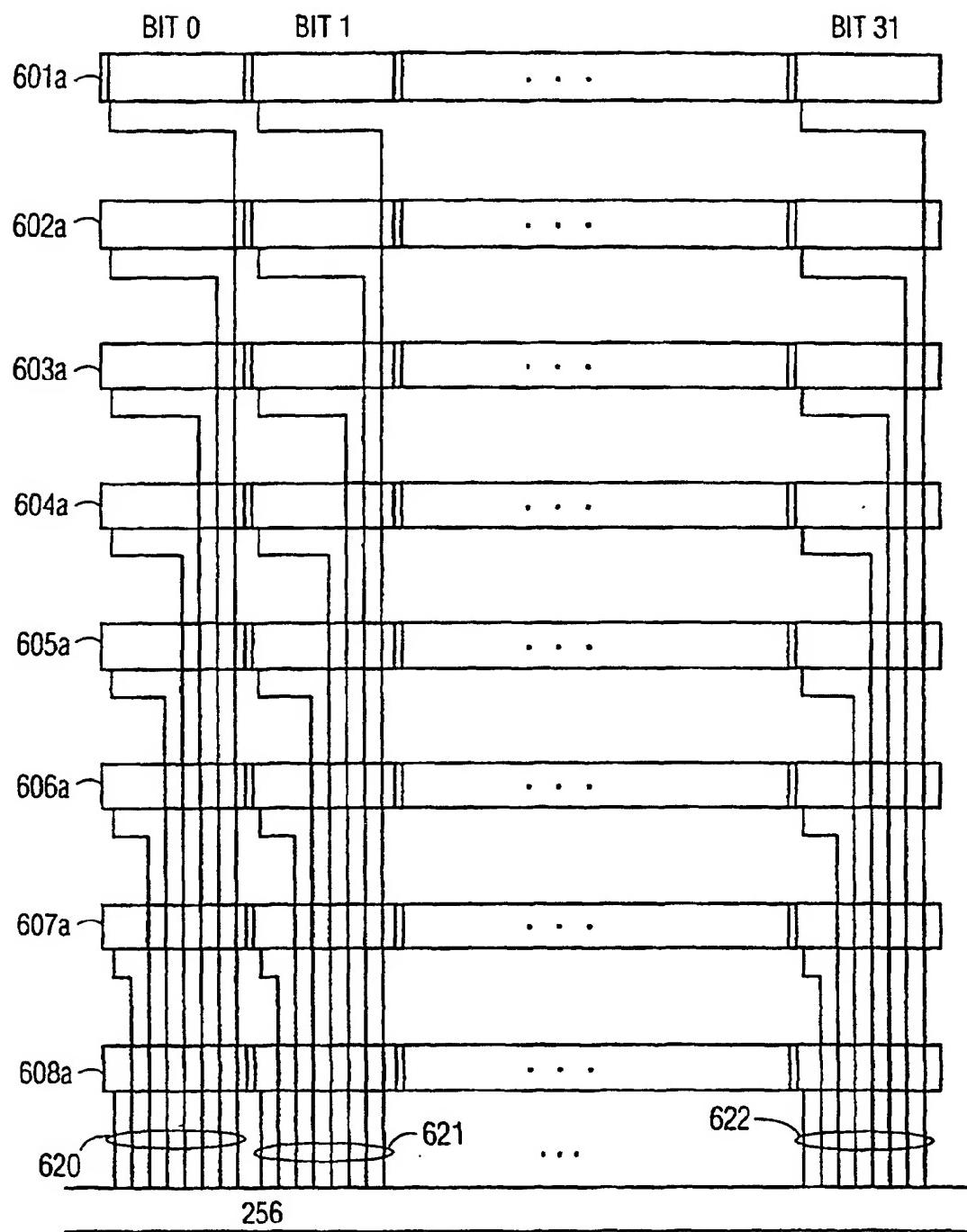


FIG. 6B

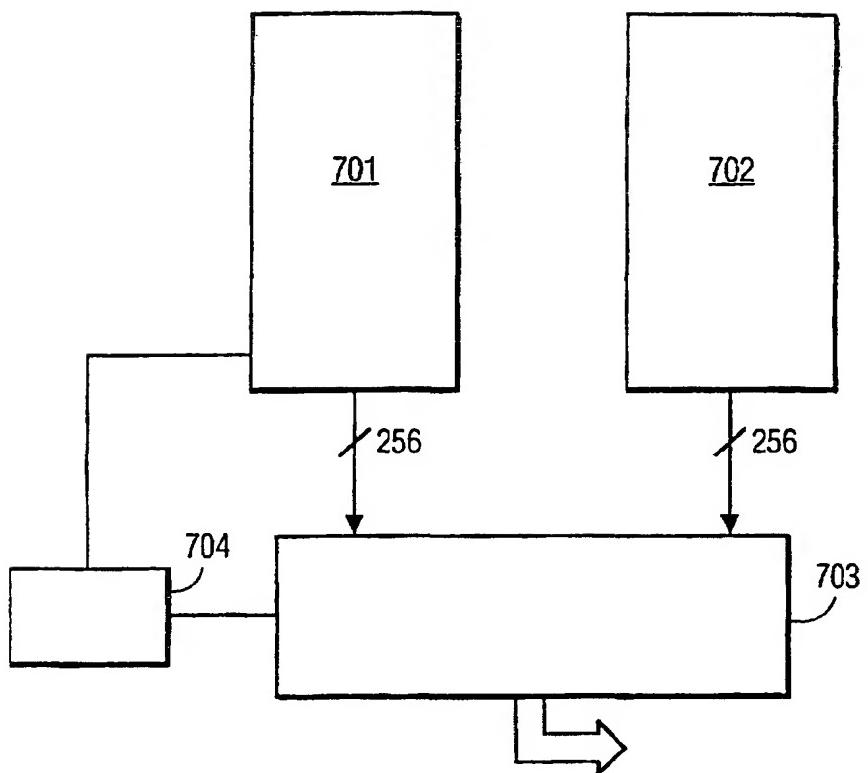


FIG. 7

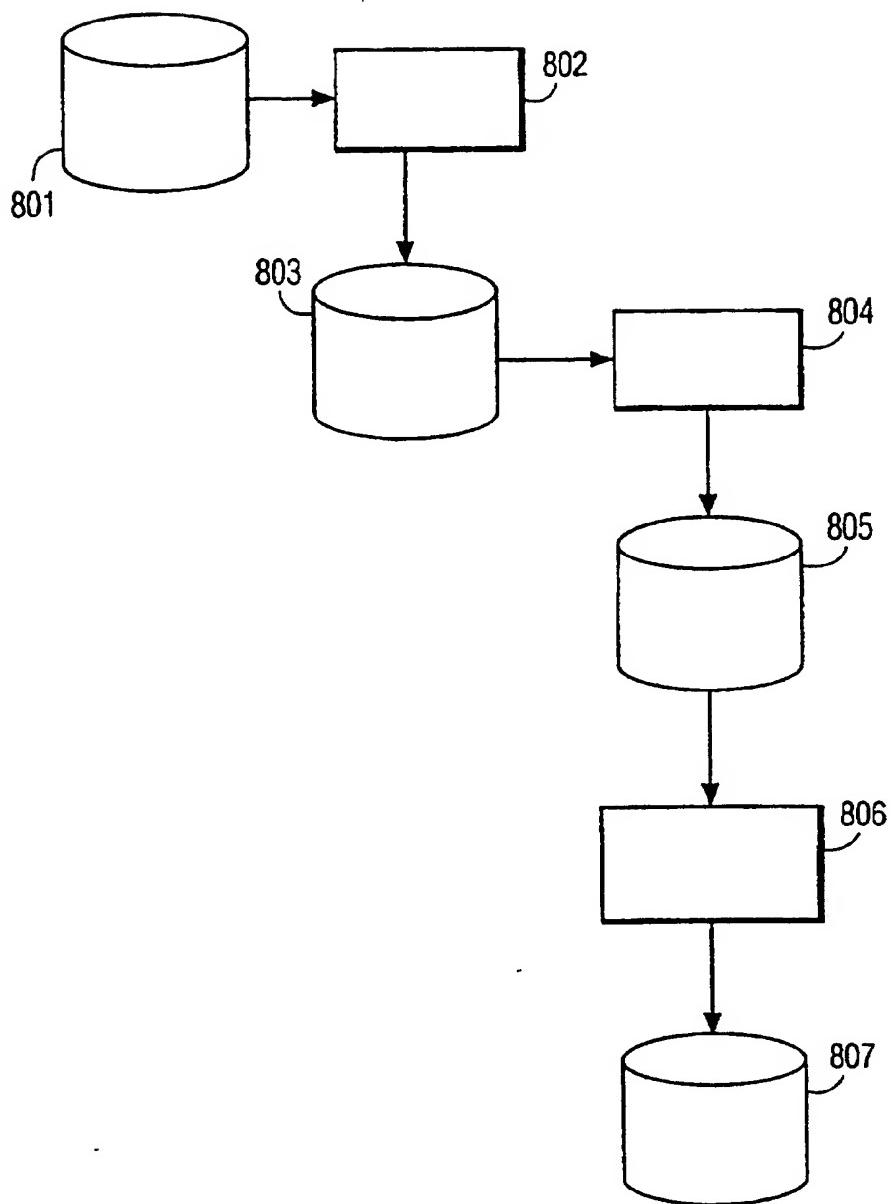


FIG. 8

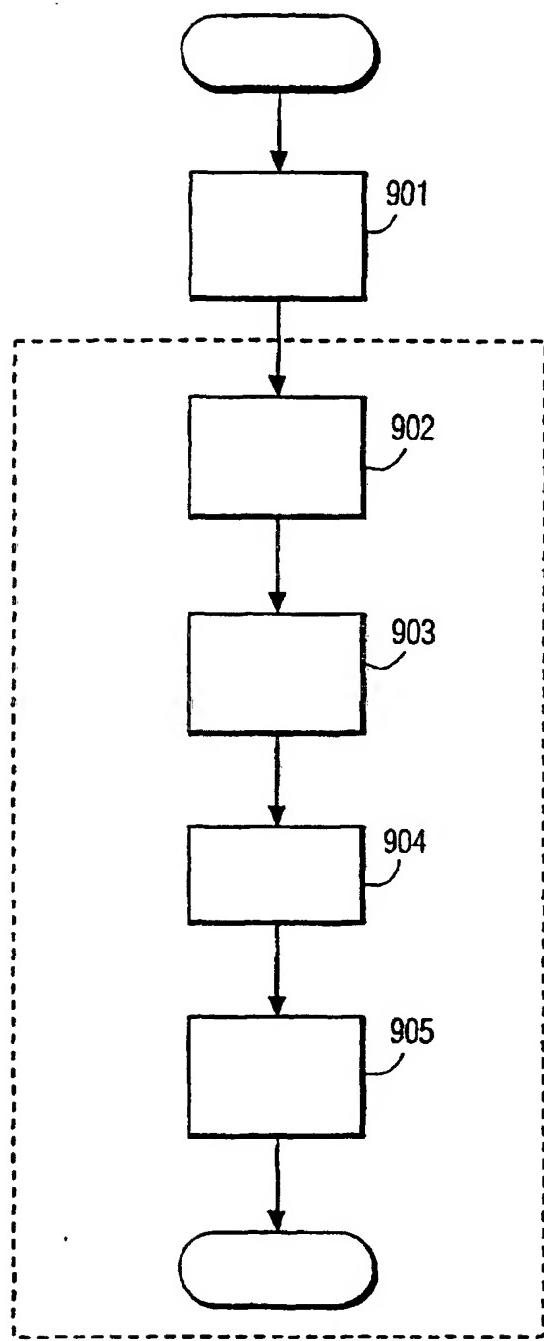


FIG. 9

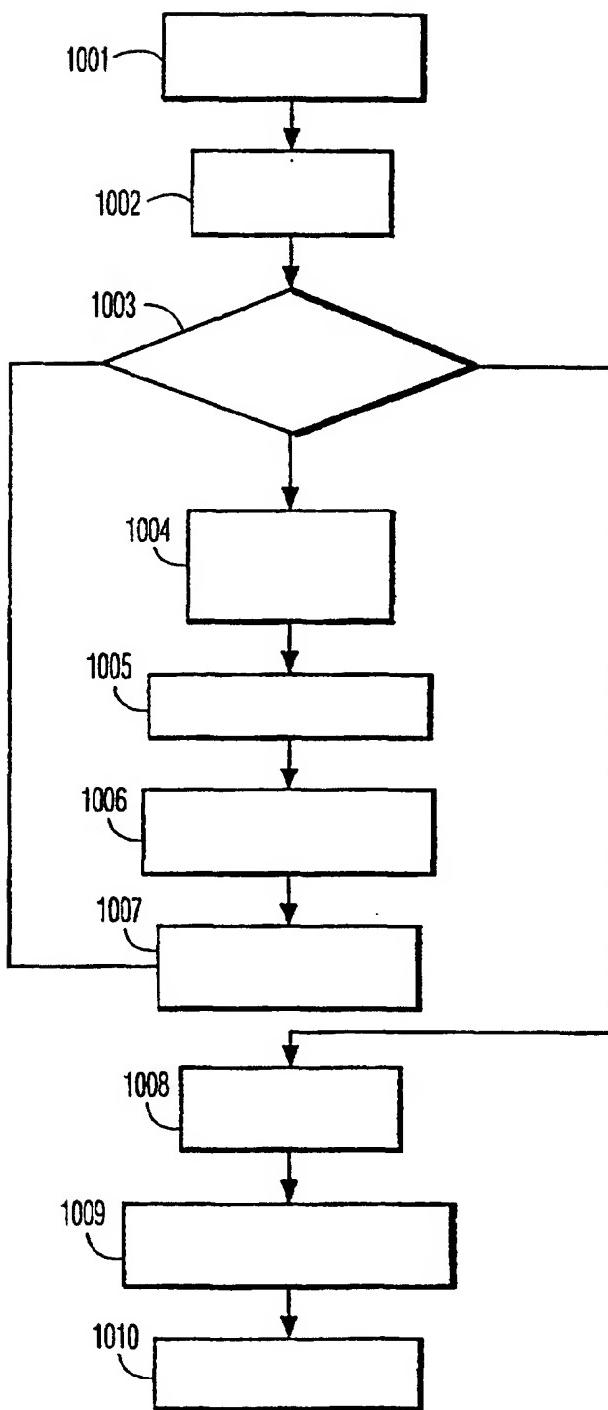


FIG. 10

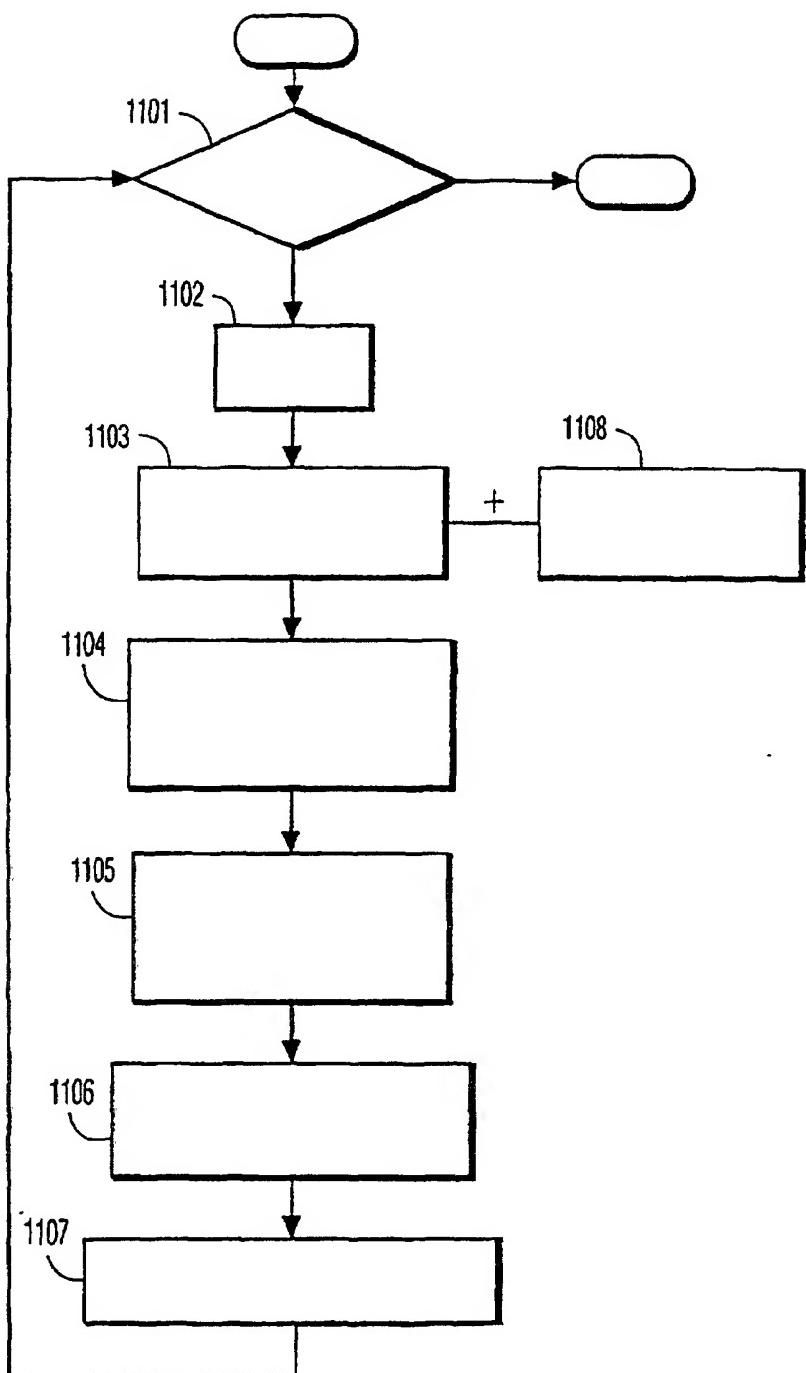


FIG. 11

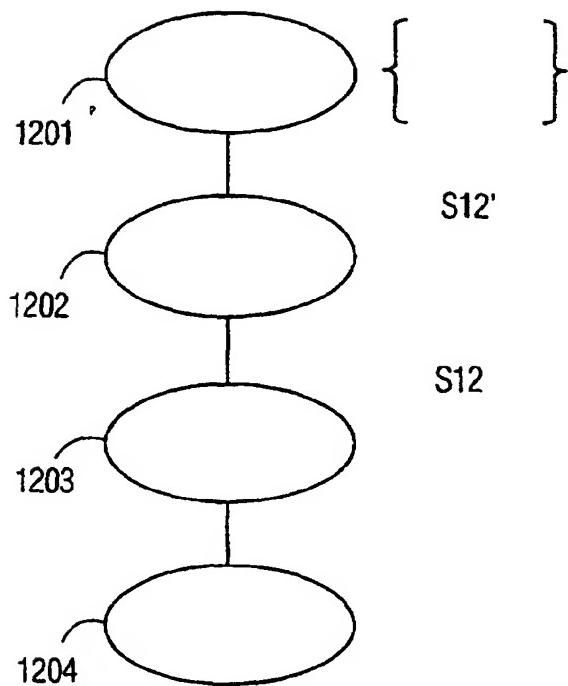


FIG. 12